

A RESOURCE-LIGHT METHOD AND APPARATUS FOR OUTLIER DETECTION

ABSTRACT OF THE DISCLOSURE

Outlier detection methods and apparatus have light computational
5 resources requirement, especially on the storage requirement, and yet achieve
a state-of-the-art predictive performance. The outlier detection problem is first
reduced to that of a classification learning problem, and then selective
sampling based on uncertainty of prediction is applied to further reduce the
amount of data required for data analysis, resulting in enhanced predictive
10 performance. The reduction to classification essentially consists in using the
unlabeled normal data as positive examples, and randomly generated
synthesized examples as negative examples. Application of selective sampling
makes use of an underlying, arbitrary classification learning algorithm, the
data labeled by the above procedure, and proceeds iteratively. Each iteration
15 consisting of selection of a smaller sub-sample from the input data, training of
the underlying classification algorithm with the selected data, and storing the
classifier output by the classification algorithm. The selection is done by
essentially choosing examples that are harder to classify with the classifiers
obtained in the preceding iterations. The final output hypothesis is a voting
20 function of the classifiers obtained in the iterations of the above procedure.